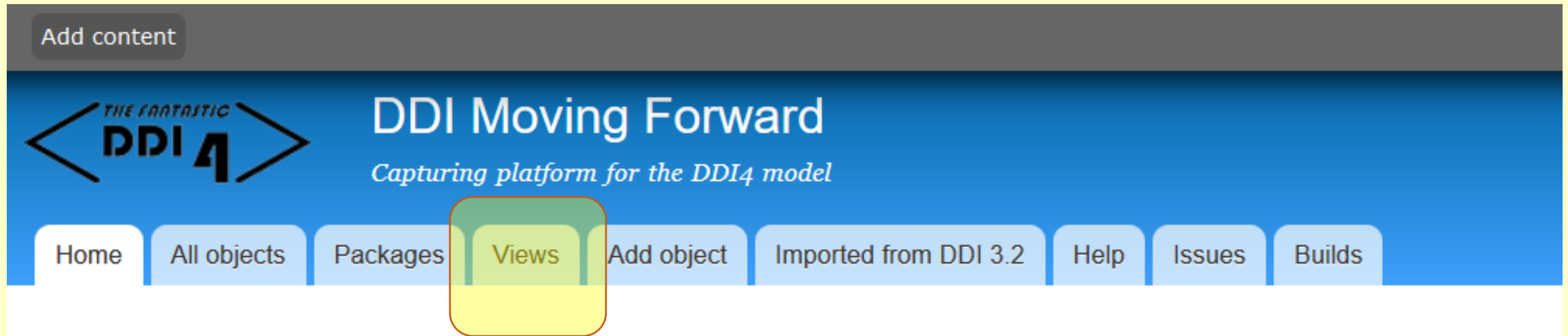


Sample Use Cases for the DataDictionary View in DDI Views (DDI4)

Dan Gillman, Arofan Gregory, Larry Hoyle, Knut Wenzig

DDI Views - <http://lion.ddialliance.org/>



DDIViews (DDI4) has **Functional Views**, which are subsets of the whole model, typically many fewer classes

All Functional Views are compatible

DataDictionary Functional View

DataDictionaryView

[View](#)[Edit](#)[What links here](#)[Revision operations](#)

Submitted by wendy on Mon, 09/19/2016 - 10:03

Purpose:

The Data Dictionary View provides the limited information on a data file structure including: the name of the data file, name of variable, physical location (order) of variable, type of variable (character, numeric, etc.), variable label, variable universe/population, substantive (valid) and sentinel (invalid) values with labels if appropriate, and variable concept.

Use Cases:

This is a core section of most codebooks as well as often being supplied as a structured document as import to various statistical package set-up files.

Target Audience:

Creators of data files for access and distributors.

Included Classes:

Access, AttributeRole, Category, CategorySet, Code, CodeItem, CodeList, Concept, Coverage DataPoint, DataRecord, DataStore, Datum, FundingInformation, IdentifierRole, Individual, InstanceVariable, InstanceVariableMapping, Level, LogicalRecordLayout, MeasureRole, Organization, PhysicalLayoutOrder, PhysicalLayoutOrderPairs, Population, RecordRelation, RectangularLayout (physicalLayout), SegmentByText, SentinelConceptualDomain, SentinelValueDomain, StructureDescription, SpatialCoverage, SubstantiveConceptualDomain, SubstantiveValueDomain, TemporalCoverage, TopicalCoverage, Unit, UnitType, Universe, ValueAndConceptDescription, ValueMapping, Viewpoint

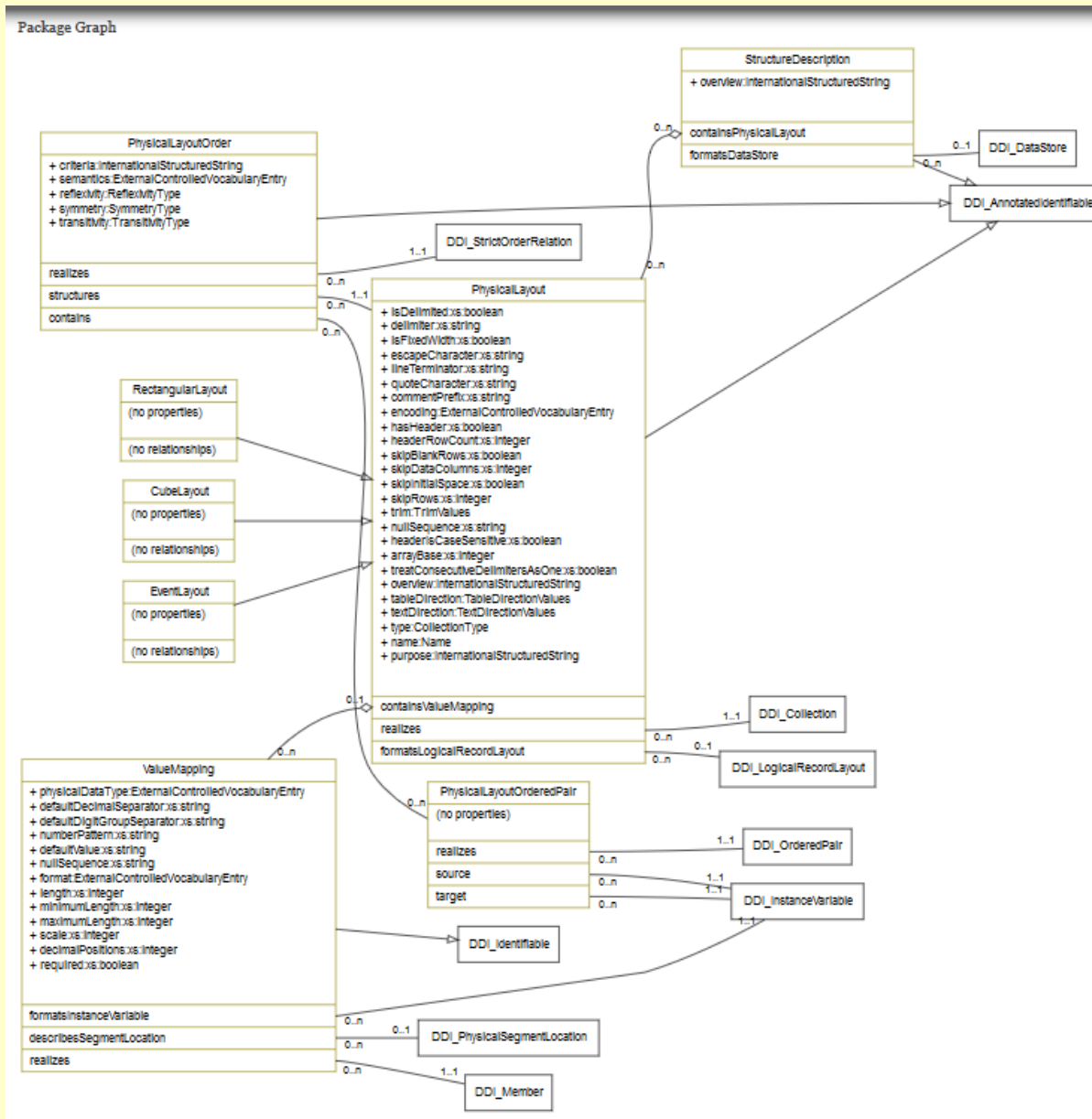
Focuses on physical file structure

Classes in the DataDictionary Functional View

- [Access](#)
- [AttributeRole](#)
- [Category](#)
- [CategorySet](#)
- [Code](#)
- [CodeItem](#)
- [CodeList](#)
- [Concept](#)
- [Coverage](#)
- [DataPoint](#)
- [DataRecord](#)
- [DataStore](#)
- [Datum](#)
- [FundingInformation](#)
- [IdentifierRole](#)
- [Individual](#)
- [InstanceVariable](#)
- [InstanceVariableMapping](#)
- [Level](#)
- [LogicalRecordLayout](#)
- [Machine](#)
- [MeasureRole](#)
- [Organization](#)
- [PhysicalLayoutOrder](#)
- [PhysicalLayoutOrderedPair](#)
- [Population](#)
- [RecordRelation](#)
- [RectangularLayout](#)
- [SegmentByText](#)
- [SentinelConceptualDomain](#)
- [SentinelValueDomain](#)
- [SpatialCoverage](#)
- [StructureDescription](#)
- [SubstantiveConceptualDomain](#)
- [SubstantiveValueDomain](#)
- [TemporalCoverage](#)
- [TopicalCoverage](#)
- [Unit](#)
- [UnitType](#)
- [Universe](#)
- [ValueAndConceptDescription](#)
- [ValueMapping](#)
- [Viewpoint](#)

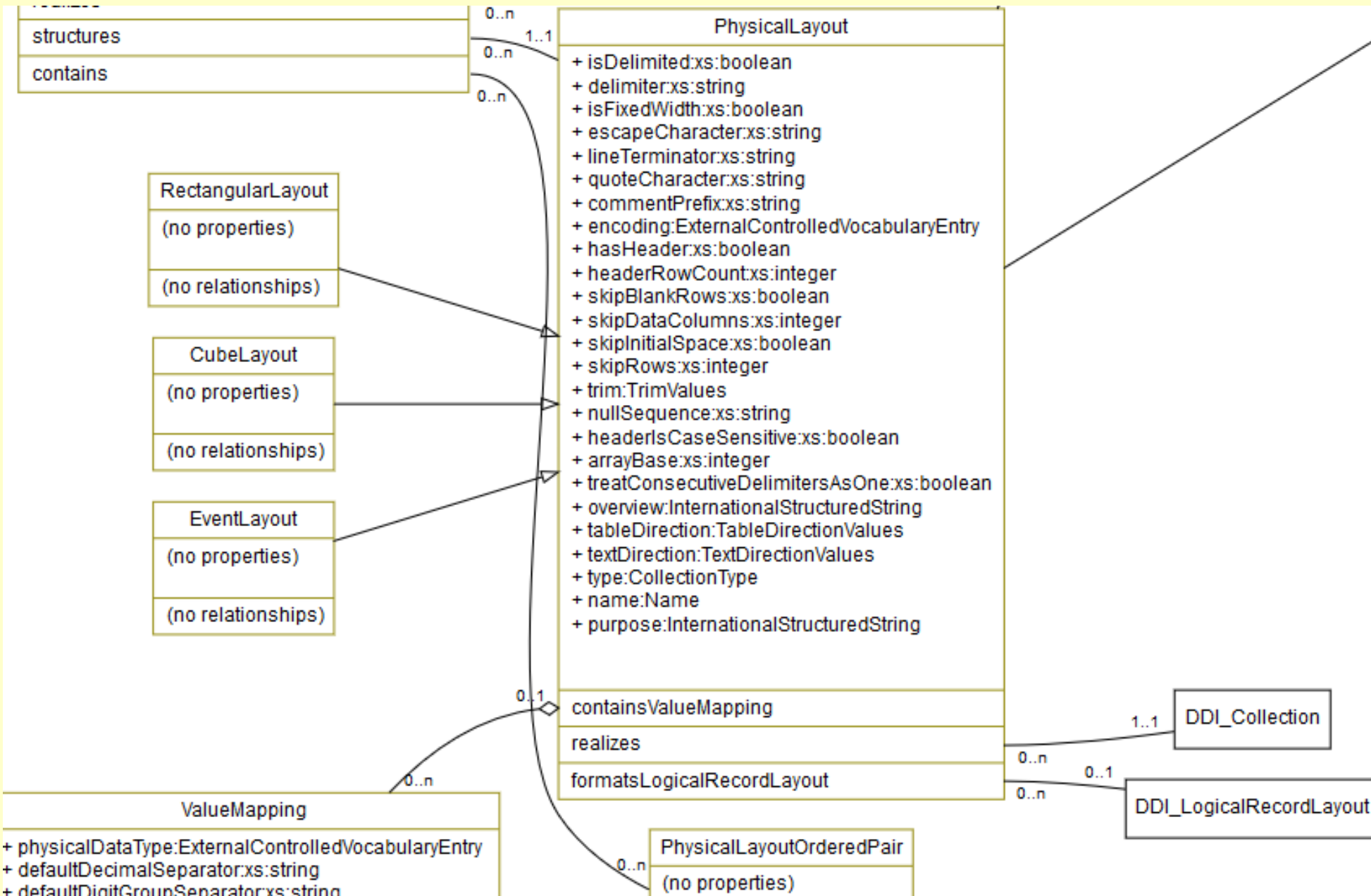
This view has 43 of the current 858 classes in Lion

DDI-Views Packages (Example: FormatDescription)



Packages are useful when looking at, or developing, the overall model. They are not namespaces.

DDI-Views Packages (Example: FormatDescription Detail)



UML model features:

- Inheritance
- Realization
- Properties
- Relationships
 - aggregations
 - plain

Describing a class in Lion – definition, notes, and examples

ValueMapping

View Edit What links here Revision operations

Submitted by larry on Wed, 09/07/2016 - 15:23

Package: [FormatDescription](#)

Extends: [Identifiable](#)

Version: 1

Is Abstract: no

Status: [Proposal in progress](#)

Contributor: [larry](#)

Definition:
Provides physical characteristics for an InstanceVariable as part of a PhysicalLayout

Example:
A variable "age" might be represented in a file as a string with a maximum length of 5 characters and a number pattern of ##0.0

Property:

Inherited from: [Identifiable](#)

Name	Cardinality	Datatype	Description
			This is the registered agency code with optional

[http://lion.ddialliance.org/
ddiobjects/valuemapping](http://lion.ddialliance.org/ddiobjects/valuemapping)

Describing a class in Lion – properties

Object properties

Name	Cardinality	Datatype	Description	Operations
physicalDataType	0..1	ExternalControlledVocabularyEntry	The base datatype of the physical representation. An integer InstanceVariable might, for example, be stored as a floating point number. From https://www.w3.org/TR/tabular-metadata/ Inherited 5.7 datatype: "An atomic property that contains either a single string that is the main datatype of the values of the cell or a datatype description object. If the value of this property is a string, it MUST be the name of one of the built-in datatypes defined in section 5.11.1 Built-in Datatypes and this value is normalized to an object whose base property is the original string value. If it is an object then it describes a more specialized datatype. If a cell contains a sequence (i.e. the separator property is specified and not null) then this property specifies the datatype of each value within that sequence. See 5.11 Datatypes and Parsing Cells in [tabular-data-model] for more details. The normalized value of this property becomes the datatype annotation for the described column. "	<ul style="list-style-type: none"> • Edit • Delete
defaultDecimalSeparator	0..1	xs:string	The string separating the integer part from the fractional part of a decimal or real number. In W3C part of the datatype format From https://www.w3.org/TR/tabular-metadata/ tabular 6.4.2 decimalChar: "A string whose value is used to represent a decimal	<ul style="list-style-type: none"> • Edit • Delete

Describing a class in Lion – relationships

Object relationships

Name	Target Object	Description	Source cardinality	Target cardinality	Relationship type	Operations
formatsInstanceVariable	InstanceVariable	Describes the physical representation of the InstanceVariable	0..n	1..1	Neither	<ul style="list-style-type: none">• Edit• Delete
describesSegmentLocation	PhysicalSegmentLocation	Uses a PhysicalSegmentLocation to describe where in the physical record a segment representing the InstanceVariable is. This could be, for example, described as a start position and end position value for characters in a text record via the SegmentByText extension of PhysicalSegmentLocation.	0..n	0..1	Neither	<ul style="list-style-type: none">• Edit• Delete
realizes	Member	ValueMappings are members of the PhysicalLayout.	0..n	1..1	Neither	<ul style="list-style-type: none">• Edit• Delete

Dagstuhl Sprint, October 2016 (Week Two)

Created by Jared Lyle, last modified by Michelle Edwards on Nov 30, 2016



Overview

The DDI (Data Documentation Initiative) metadata standard, originally created in 1995 to document social science research data, has in recent years become

Papers and Output

Preparing for the
In preparation, ple

Read These

Detailed outcom

Introduction to t
Description Pack

Atlassian Tool Us
document may b

Data Description working group

Tested the DataDescription View Physical file layouts:

- CSV Simple rectangular
- Fixed Column rectangular
- CSV Segmented
- CSV Hierarchical
- CSV Aggregate (ncube)

We also looked at event history data and think it will be describable with these elements too

see: <https://ddi-alliance.atlassian.net/wiki/display/DDI4/Data+Description+View+Team>

Event History Data

- The event is the unit of observation
- Examples
 - A person visits a doctor: datasets holds all persons' id, date of visit and type of doctor; each row is a doctor's visit of a person
 - A petrol station registers each car in case of refilling: dataset holds license plate number, date/time of refilling and amount of fuel
- Additions to DDI needed for event data
 - Which variable holds temporal information, which variable holds id of object with multiple events (see Viewpoint)
 - A datatype of “pointer” or “reference”, the value of which is a DDI InstanceVariable or perhaps a ValueMapping

Simple Rectangular CSV

<https://ddi-alliance.atlassian.net/wiki/display/DDI4/Data+Description+View+Team>

DDI Example of Use: Unit Record Data in CSV

Contributors: Daniel Gillman (Bureau of Labor Statistics), Arofan Gregory (Aeon Technologies), Larry Hoyle (University of Kansas, Institute for Policy and Social Research), Knut Wenzig (DIW Berlin)

A. Business Case

The Australian Election Survey for 2013 is encoded in a comma-delimited format. The Australian Data Archive provides access to this data set. It is described in two ways – as a minimal data description to support processing of the data, and in a more complete fashion. Both are provided in this example.

B. Relevant Classes from the Model

In this section we will apply the Data Dictionary view, as it combines both logical and physical description. Metadata which is not related to the description of the data, but instead is related to the study is not covered.

Item	DDI 4 Construct	Notes
Variable name	InstanceVariable.name	



More details are available on the web in the report from the working group

Simple Rectangular CSV – Australian Election Survey

```
0          10          20          30          40          50          60          70          80
1 divisnum,uniqueid,mode,datecomp,state,division,a4,glage,xg5,weight,partyaby,swingn
2 101,3203991,2,10/22/2013,8,CANBERRA,3,18,2342,1.40422425242532,ALP,2.17000007629394
3 101,3210858,2,10/25/2013,8,CANBERRA,1,17,2241,.851715011726785,ALP,2.17000007629394
4 101,3336161,1,9/16/2013,8,CANBERRA,1,17,1325,.911903409714549,ALP,2.17000007629394
5 101,3524779,1,10/4/2013,8,CANBERRA,1,15,5211,.77475821048464,ALP,2.17000007629394
6 101,3527491,1,10/28/2013,8,CANBERRA,1,17,2114,.982439220788559,ALP,2.17000007629394
7 101,3685033,1,9/19/2013,8,CANBERRA,2,18,4233,1.91487075184003,ALP,2.17000007629394
8 101,4080595,1,9/23/2013,8,CANBERRA,2,17,5421,1.01078853174099,ALP,2.17000007629394
9 101,4176643,1,9/23/2013,8,CANBERRA,2,17,2244,6.46615621820264,ALP,2.17000007629394
10 101,4302622,1,9/23/2013,8,CANBERRA,1,17,5111,.79784558088334,ALP,2.17000007629394
11 101,4766943,1,10/15/2013,8,CANBERRA,3,16,4233,.77475821048464,ALP,2.17000007629394
```

The example uses a small
representative set of variables

See: <http://www.australianelectionstudy.org/about.html>

Example DDI4 XML

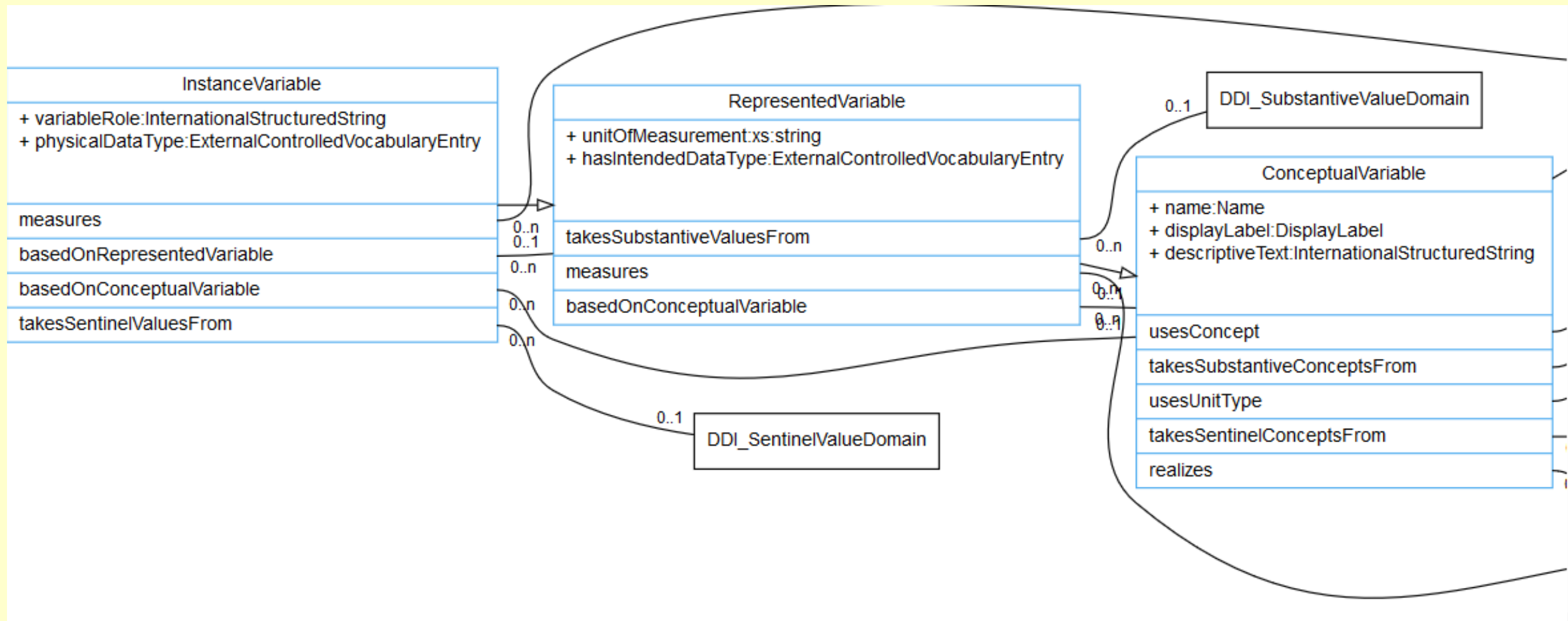
E. XML Example

```
<?xml version="1.0" encoding="UTF-8"?>  
<DDI xmlns="urn:ddi.org:4"  
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  xsi:schemaLocation="urn:ddi.org:4 file:/C:/Dagstuhl/ddi4_2016-10-25/xsd/DataDictionaryView_4-DR0.2.xsd"  
  type="DataDictionaryView">  
  <DocumentInformation>  
    <Agency>dagstuhl16433.ddialliance.org</Agency>  
    <Id>DagDocinfoTest1</Id>  
    <Version>1</Version>  
    <VersionDate>2016-10-24</VersionDate>
```

More XML

```
<Contributor>
  <Agent>
    <String>Dan Gillman</String>
    <Affiliation>BLS - U.S. Bureau of Labor Statistics</Affiliation>
  </Agent>
  <Role>
    <ControlledVocabularyAgencyName>casrai</ControlledVocabularyAgencyName>
    <ControlledVocabularyName>Contributor Roles</ControlledVocabularyName>
    <Content>Conceptualization</Content>
    <Extent>Equal</Extent>
  </Role>
</Contributor>
```

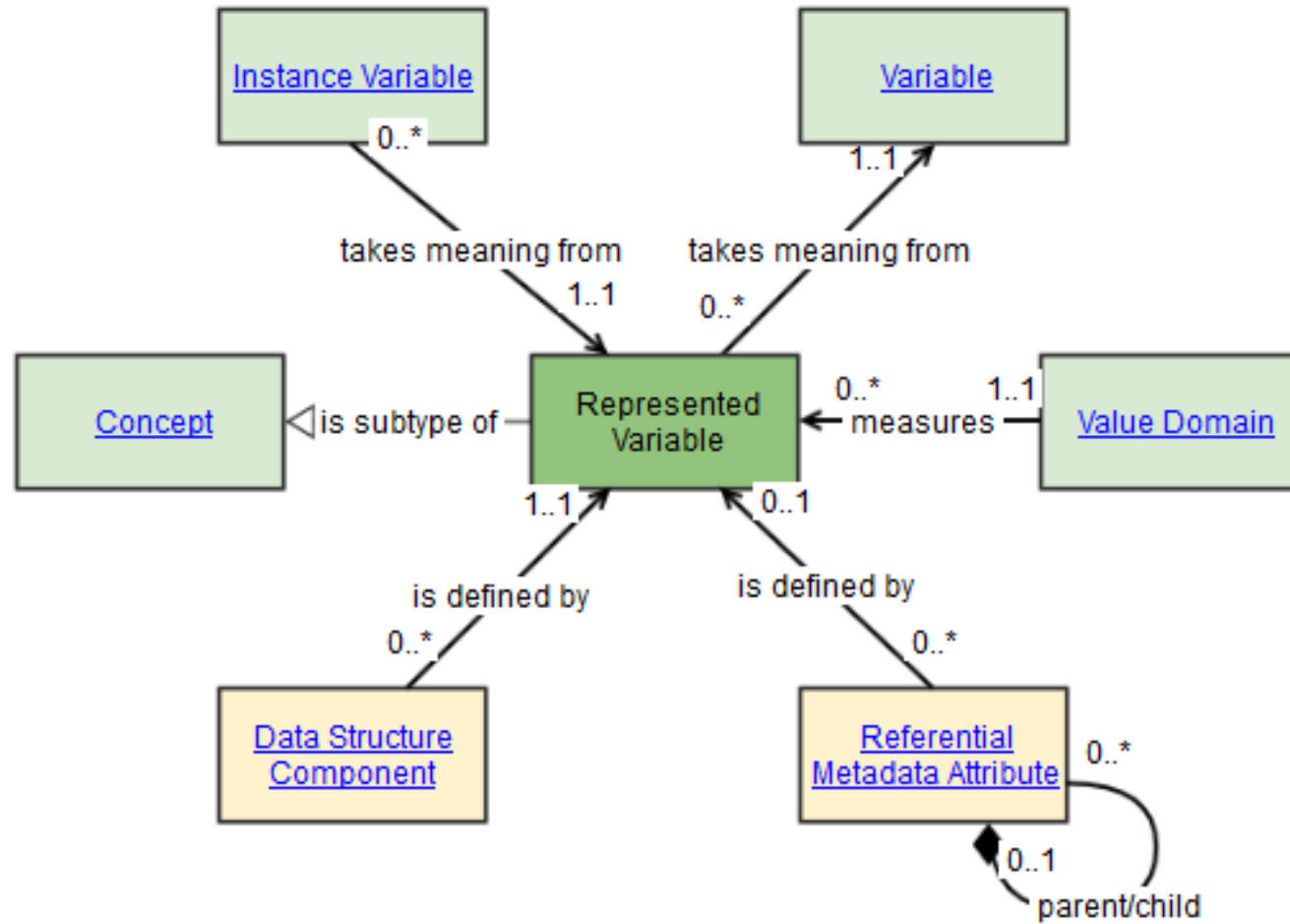
Variable Cascade in DDI <http://lion.ddialliance.org/package/conceptual>



The InstanceVariable can reference RepresentedVariable and ConceptualVariable
It also inherits all properties and relationships from them.

Modeled to comply with GSIM

<http://www1.unece.org/stat/platform/display/GSIMclick/Instance+Variable>
<http://www1.unece.org/stat/platform/display/GSIMclick/Represented+Variable>



Logical and physical records – at the record level

Logical

- Variables
- Logical variable order

Physical

- LogicalRecord link
- Physical variable order
- Number physical records per logical
- Encoding etc.
- Record parameters (e.g. terminator)
- Delimited?/delimiter, quote char ...

Physical Records – at the variable level

- Physical representation
 - Datatype, decimal separator, group separator ...
- Location
 - Starting position, ending position, length
 - Record number

Describing a CSV

Logical

InstanceVariable

LogicalRecordLayout

```
1 divisnum,uniqueid,mode,datecomp,state,division,a4,glage,xg5,weight,partyaby,swingn  
2 101,3203991,2,10/22/2013,8,CANBERRA,3,18,2342,1.40422425242532,ALP,2.17000007629394
```

ValueMapping

RectangularLayout

PhysicalLayoutOrder

Physical

Relationships

Instance Variable - Properties And Relationships

<http://lion.ddialliance.org/ddiobjects/instancevariable>

- Agency
- ID
- Version
- Name
- DisplayLabel
- DescriptiveText
- [ANNOTATION](#)
- UnitOfMeasurement
- VariableRole
- PhysicalDatatype
- IntendedDatatype
- SubstantiveConceptualDomain
- SubstantiveValueDomain
- SentinelConceptualDomain
- SentinelValueDomain
- Concept
- UnitType
- Universe
- Population
- ConceptualVariable
- RepresentedVariable
- ExternalMaterial

LogicalRecordLayout - Properties and Relationships <http://lion.ddialliance.org/ddiobjects/logicalrecordlayout>

- Agency, Id, Version
- Annotation
- Type
- Name
- Purpose
- **InstanceVariable**
- ViewPoint
- LogicalReordLayout
- ExternalMaterial
- **Realizes Collection**

As a Collection the LogicalRecordLayout allows for ordering InstanceVariables within the Record.

Viewpoints allow assigning roles to InstanceVariables – Identifiers, Measures, or Attributes.

RectangularLayout (PhysicalLayout) – Properties and Relationships <http://lion.ddialliance.org/ddiobjects/physicallayout>

- Agency, ID, Version
- Annotation
- IsDelimited
- Delimiter
- IsFixedWidth
- EscapeCharacter
- LineTerminator
- QuoteCharacter
- CommentPrefix
- Encoding
- Header
- HeaderRowCount
- SkipBankRows
- SkipDataColumns
- SkipInitialSpace
- SkipRows
- Trim
- NullSequence
- HeaderIsCaseSensitive
- ArrayBase
- TreatConsecutiveDelimitersAs One
- Overview
- TableDirection
- TextDirection
- Type
- Name
- Purpose
- ExternalMaterial
- **ValueMapping**
- **LogicalRecordLayout**
- Realizes Collection

ValueMapping - Properties and Relationships <http://lion.ddialliance.org/ddiobjects/valuemapping>

- Agency, Id, Version
- PhysicalDatatype
- DefaultDecimalSeparator
- DefaultDigitalGroupSeparator
- NumberPattern
- DefaultValue
- NullSequence
- Format
- Length
- MinimumLength
- MaximumLength
- Scale
- DecimalPositions
- Required
- ExternalMaterial
- **InstanceVariable**
- PhysicalSegmentLocation
 - (e.g. Start, end, length)

PhysicalLayoutOrder - Properties and Relationships <http://lion.ddialliance.org/ddiobjects/valuemapping>

Order Criteria (alphabetical, numerical ...)

Semantics (before)

Reflexivity (Anti-Reflexive – a variable is not before itself)

Symmetry (Anti – A before B means B not before A)

Transitivity(

Transitive (if A before B and B before C then A before C)

PhysicalLayout it structures

PhysicalLayoutOrderedPairs

PhysicalLayoutOrderedPair - Properties and Relationships <http://lion.ddialliance.org/ddiobjects/valuemapping>

Source InstanceVariable

Target InstanceVariable

The source variable is “before” the target variable.

The PhysicalLayoutOrder defined the semantic: “before”

Example:

divisnum before uniqueid

uniqueid before mode

...

Typical CSV RectangularLayout in XML

```
<RectangularLayout>  
  <Agency>dagstuhl16433.ddialliance.org</Agency>  
  <Id>L12345</Id>  
  <Version>1</Version>  
  <Encoding>ASCII</Encoding>  
  <HasHeader>>true</HasHeader>  
  <HeaderRowCount>1</HeaderRowCount>  
  <IsDelimited>true</IsDelimited>  
  <Delimiter>,</Delimiter>  
  <LineTerminator>#xD#xA</LineTerminator>  
  <ContainsValueMapping typeOfClass="ValueMapping"  
URI="URN:DDI:dagstuhl16433.ddialliance.org:VM_DivisNumID:1">  
  </ContainsValueMapping>  
  <ContainsValueMapping typeOfClass="ValueMapping"  
URI="URN:DDI:dagstuhl16433.ddialliance.org:VM_UniqueIDID:1">  
  </ContainsValueMapping>
```

InstanceVariable for variable “DivisNum” in XML

```
<InstanceVariable>  
  <Agency>dagstuhl16433.ddialliance.org</Agency>  
  <Id>DivisNumID</Id>  
  <Version>1</Version>  
  <Name><Content>DivisNum</Content></Name>  
  <DisplayLabel><Content>Electoral Division – AES Numeric Code</Content></DisplayLabel>  
  <HasIntendedDataType>Nominal</HasIntendedDataType>  
</InstanceVariable>
```

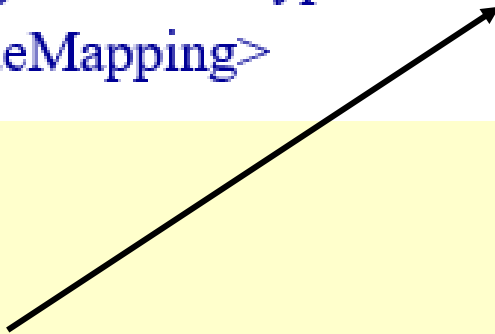
Actually an
“ExternalControlledVocabularyEntry”



Much more about the ValueDomain for
DivisNum could be described with
SubstantiveValueDomain and
SentinelValueDomain

CSV ValueMapping for variable “DivisNum” in XML

```
<ValueMapping>
  <Agency>dagstuhl16433.ddialliance.org</Agency>
  <Id>VM_DivisNumID</Id>
  <Version>1</Version>
  <FormatsInstanceVariable typeOfClass="InstanceVariable" isExternal="false"
  URI="URN:DDI:dagstuhl16433.ddialliance.org:DivisNumID:1" ></FormatsInstanceVariable>
  <PhysicalDataType>numeric code</PhysicalDataType>
</ValueMapping>
```



Also an “ExternalControlledVocabularyEntry”

Fixed Layout ValueMapping for variable "DivisNum" in XML

```
<ValueMapping>
  <Agency>dagstuhl16433.ddialliance.org</Agency>
  <Id>VM_DivisNumID</Id>
  <Version>1</Version>
  <FormatsInstanceVariable typeOfClass="InstanceVariable" isExternal="false"
URI="URN:DDI:dagstuhl16433.ddialliance.org:DivisNumID:1" ></FormatsInstanceVariable>
  <PhysicalDataType>numeric code</PhysicalDataType>
  <DescribesSegmentLocation typeOfClass="SegmentByText"
URI="URN:DDI:dagstuhl16433.ddialliance.org:SBT_DivisNumID:1"></DescribesSegmentLocation>
</ValueMapping>
<SegmentByText>
  <Agency>dagstuhl16433.ddialliance.org</Agency>
  <Id>SBT_DivisNumID</Id>
  <Version>1</Version>
  <DefinedByLineParameters>
    <StartLine>1</StartLine>
    <StartOffset>1</StartOffset>
    <EndLine>1</EndLine>
    <EndOffset>3</EndOffset>
  </DefinedByLineParameters>
</SegmentByText>
```

Viewpoint <http://lion.ddialliance.org/ddiobjects/viewpoint>

- Assigns one of three roles to variables:
 - Identifier
 - Measure
 - Attribute
- Useful for
 - Aggregate data (ncubes)
 - Event data (Which variable holds temporal information, which variable holds id of object with multiple events)
 - Annotating variables (e.g. paradata)